

Обзор генома и протеома бактерии *Streptomyces fodineus*

Автор: М.С. Давитадзе

РЕЗЮМЕ

В данном обзоре я провела анализ нуклеотидной последовательности генома *Streptomyces fodineus*, проанализировала разнообразие генов в геноме и их местоположение. Данные, полученные мной в этой работе, соответствуют данным о геноме *Streptomyces fodineus* из других источников.

СПИСОК СОКРАЩЕНИЙ

SRP-частица - signal recognition particle, частицы распознавания сигнала

1. ВВЕДЕНИЕ

Бактерии рода *Streptomyces* относятся к актинобактериям. Это грамположительные аэробные нитевидные бактерии. Для них характерны следующие признаки: ДНК с повышенным содержанием гуанина и цитозина, жирные кислоты с сильно разветвленными цепочками и наличие LL-диаминопимелиновой кислоты в клеточной стенке. Эти бактерии образуют обильно ветвящийся мицелий и цепочки спор. Они являются основными представителями почвенных бактерий. Штамм *Streptomyces fodineus* был выделен из образца кислой почвы, взятого возле шахты. Эти бактерии образуют бело-серый или желтовато-коричневый воздушный мицелий и серо-белый субстратный мицелий, споры их гладкие. Они также продуцируют противогрибковые вещества. *Streptomyces fodineus* устойчивы к пенициллину, ампициллину и хлорамфениколу [1].

Цели и задачи

В данной работе я поставила перед собой следующие цели: проанализировать нуклеотидную последовательность генома *Streptomyces fodineus*, выяснить, какие гены в нем закодированы, а также проанализировать гены, кодирующие белки.

2. МАТЕРИАЛЫ И МЕТОДЫ

Для обработки данных я пользовалась электронными таблицами "Google таблицы", языком программирования Python 3, а также командной оболочкой Bash.

Анализ нуклеотидной последовательности ДНК

С помощью написанного мной на Python скрипта (сопроводительные материалы, пункт 1), я вычислила длину нуклеотидной последовательности генома. Также с помощью Bash я посчитала число нуклеотидов в геноме (сопроводительные материалы, пункт 2) и количество димеров (сопроводительные материалы, пункт 3) каждого типа, а также их процент от общего числа. Я вычислила GC-состав, используя формулу:

$$GC(\%) = ((G + C) \div L) \cdot 100$$

где L - это число всех нуклеотидов одной цепи ДНК: $A + T + G + C$ (табл.1).

Анализ генов

С помощью функции СЧЁТЕСЛИ Google таблиц я нашла число генов разных типов и посчитала их общее количество, а также нашла процент каждого типа от общего числа генов. Также с помощью функции СЧЁТЕСЛИМН я вычислила процент генов, закодированных на + и - цепях ДНК.

Анализ длин белков

С помощью функции Google таблиц ЕСЛИ я составила таблицу длин белков *Streptomyces fodineus* в аминокислотных остатках. На ее основе я построила гистограмму. Также с помощью функций МАКС и МИН я нашла самый длинный и самый короткий белок, среднюю длину белков (с помощью функции СРЗНАЧ) и медианное значение длин белков (функция МЕДИАНА).

3. РЕЗУЛЬТАТЫ

Анализ нуклеотидной последовательности ДНК

Длина генома *Streptomyces fodineus* - 9698948 пар нуклеотидов. В геномной последовательности встречаются только буквы А, Т, G и С - нуклеотиды аденин, тимин, гуанин и цитозин. При этом число аденинов примерно равно числу тиминов, а число гуанинов примерно равно числу цитозинов (табл.1).

Табл.1 количество нуклеотидов каждого типа (в шт. нуклеотидов)

тип нуклеотида	количество
Аденин	1408226
Тимин	1395856
Гуанин	3443047
Цитозин	3451819

Гуанин и цитозин составляют 71% от общего числа нуклеотидов ДНК. Также оказалось, что наиболее часто в геноме встречаются димеры CG (14%), а самые редкие димеры - TA (1%) (табл.2).

Табл.2 количество и процент димеров каждого типа

2-mers	count	percent
CG	1373701	14
GC	1205512	12
CC	1084175	11
GG	1075713	11
GA	610179	6
TC	608558	6
AC	553574	6
GT	551643	6
CA	525288	5
TG	515103	5
AG	478529	5
CT	468654	5
AT	208007	2
AA	168116	2
TT	167552	2
TA	104643	1

Анализ генов

Всего в геноме *Streptomyces fodineus* 8714 генов, при этом большую их часть составляют гены, кодирующие белки (92%). Также в геноме присутствуют псевдогены, составляющие 7% от общего числа генов. Генов же, кодирующих РНК, в геноме всего 1,03% (в т.ч. 0,79% тРНК и 0,21% рРНК) (табл.3).

Табл.3 количество генов каждого типа и их процент от общего числа генов

тип гена	количество	процент
proteins	7997	91,77
pseudogenes	627	7,2
tRna	69	0,79
rRna	18	0,21
tmRNA	1	0,01
SRP_RNA	1	0,01
RNase_P_RNA	1	0,01
ncRNA	0	0
antisense_RNA	0	0

Выяснилось, что на - цепи содержится в целом больше генов (52%), чем на + цепи. При этом на + цепи больше генов, кодирующих рРНК, а также есть гены тмРНК, SRP-частицы и РНКазы Р, которых нет на - цепи (табл.4).

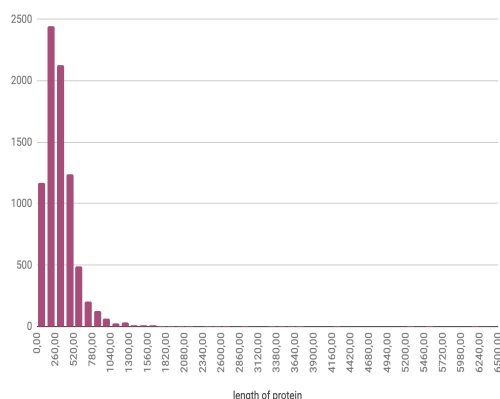
Табл.4 распределение генов по цепям ДНК

gene type	+ strand	- strand
protein coding genes	3878	4119
pseudogenes	300	327
tRna genes	21	48
rRna genes	12	6
tmRNA genes	1	0
SRP_RNA genes	1	0
RNase_P_RNA genes	1	0

Анализ длин белков

В среднем длина белков составляет 328 аминокислотных остатков, медианное значение при этом равно 278. Самый длинный белок *Streptomyces fodineus* состоит из 6145 аминокислотных остатков, а самый короткий - из 17 (рис.1, сопроводительные материалы, таблица protein_length).

Рис.1 гистограмма длин белков



4. ОБСУЖДЕНИЕ

Результаты, полученные мной в результате работы, совпали с уже имеющимися данными о геноме *Streptomyces fodineus*: геном этих бактерий состоит из 9698948 пар нуклеотидов, в нем всего 8714 генов, 7997 из которых кодируют белки [5]. В результате проведения исследования я выяснила, что в последовательности одной цепочки геномной ДНК *Streptomyces fodineus* число аденинов примерно равно числу тиминнов, а число гуанинов приблизительно равно числу цитозинов, а это значит, что для генома этой бактерии выполняется второе правило Чаргаффа. В ходе работы я также выявила, что доля гуанина и цитозина в геноме - 71%, что совпадает со значением GC-состава ДНК *Streptomyces fodineus*, полученным в других исследованиях [1]. Вероятно, такой высокий процент этих нуклеотидов связан с тем, что для данного штамма важно, чтобы ДНК была более устойчива. Гуанин соединен с цитозином тремя водородными связями, в то время как аденин с тимином - двумя, поэтому высокая доля GC-пар делает ДНК более стабильной.

5. ЗАКЛЮЧЕНИЕ

В результате работы мне удалось проанализировать нуклеотидный состав генома и найти соотношение нуклеотидных димеров в ДНК *Streptomyces fodineus*. Также я провела анализ генов, закодированных в геноме и выяснила, какие типы генов в нем присутствуют, а также выявила число генов каждого типа. Кроме того, я посчитала длину генов, кодирующих белки.

БЛАГОДАРНОСТИ

Хочу сказать огромное спасибо Андрею Владимировичу Алексеевскому, который учил меня работать с таблицами и наставлял при выполнении данного обзора. Спасибо Андрею Владимировичу за идею проведения подобного исследования. Выражаю благодарность Сергею Александровичу Спирину, Ивану Русинову и всем остальным преподавателям биоинформатики ФББ, благодаря которым я смогла освоить Bash и Python. Без них эта работа также не могла быть выполнена. И, конечно, благодарю своих однокурсников, подавших идеи для исследования генома, которые были частично реализованы в данной работе.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Документ со скриптами и командами:

<https://docs.google.com/document/d/1U6ZJOzQ-7l5UnlePoDfj171kCNKrt08ejJVGx-peZQ/edit?usp=sharing>

- 1) скрипт для вычисления длины последовательности, написанный на Python;
- 2) команда Bash для подсчета числа нуклеотидов каждого типа;
- 3) команда Bash для подсчета числа нуклеотидных димеров каждого типа;
- 4) список использованных мной функций и методов в Google таблицах.

Ссылка на таблицу protein_length с длинами белков и гистограмму длин этих белков https://docs.google.com/spreadsheets/d/1jDNa79VRHxwNYkK6Z7-bCiH_w9atdRrOffeOf5KLxhk/edit?usp=sharing

СПИСОК ЛИТЕРАТУРЫ

Min-Kyeong Kim, Hye-Jeong Kang, Su Gwon Roh, Ji Sun Park, Seung Bum Kim, *Streptomyces fodineus* sp. nov., an actinobacterium with antifungal activity isolated from mine area soil, International Journal of Systematic and Evolutionary Microbiology 69:1350–1354, 2019.

Директория с данными о геноме *Streptomyces fodineus* на сайте NCBI ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/735/805/GCF_001735805.1_ASM173580v1

Страница сайта NCBI со сведениями о геноме *Streptomyces fodineus* https://www.ncbi.nlm.nih.gov/assembly/GCF_001735805